**Project title: Prose rhythm and linguistic theory**

The goal is to automatically detect, visualize, and analyze rhythmic variation in prose texts.

1. Concept and goals

Languages are variable at all levels. For example, the two sentences *He gave names to them* and *He gave them names* convey the same content, but differ in word order, and crucially, in rhythm. In a landmark study, Liberman (1975) suggests that all temporally ordered behavior is rhythmically organized, including speech, music, and dance. This is particularly clear in songs and poetry where the natural rhythm of speech is set against a fixed conventionalized meter, creating a tension that can be manipulated for aesthetic effect. However, all speech and writing is rhythmically structured in ways that vary across individuals, styles, and genres. In linguistics such variation has often been disregarded as noise. This is reflected in several fundamental distinctions, such as langue vs. parole (de Saussure) and competence vs. performance (Chomsky). The goal of this project is to show that rhythmic variation plays an important role in shaping language structure through the study of rhythm in non-metrical prose texts (Anttila 1997).

In order to study rhythmic variation on a large scale and to make the research results useful for the larger community, we will need to develop computational tools for three purposes:

- **Metrical Parsing**: Detecting rhythmic patterns in texts based on linguistic criteria
- **Rhythm Visualization**: Visually displaying rhythm in an easy-to-understand fashion
- **Linguistic Analysis**: Modeling rhythmic variation in linguistic terms

Our project builds on modern linguistics, but draws on the rhetorical tradition that has been largely neglected by modern linguists. The ancients believed that sentences were pervaded by rhythm, or “number”, and that the ends of periods, cola, and commata were rhythmically most important (Croll 1919, Tempest 1930), something that may be related to the principle of closure in verse. One question of particular interest for this tradition was how to end a sentence successfully. Several cadences were identified (e.g., *cursus planus, cursus tardus, cursus velox*). One of our goals is to explore the cadential structure of prose with the help of computational tools and to make real progress in this almost completely neglected field. We will focus on English and Finnish for several reasons: both languages exhibit extensive rhythmic variation; we have corpora and annotation software available for both languages; and we know both languages well.

The results of our project go far beyond linguistics and will be of interest for anyone interested in prose rhythm for whatever purpose. Our goal is to produce a web application that would allow anyone to paste in texts, analyze them rhythmically, and visualize the results. One can ask, for example: Is this text rhythmically monotonous or variable? Are there clumsy sentences? Is it prose or verse? How do the two differ anyway? How does a text by author A differ from a text by author B? How do the cadences in the campaign speeches of presidential candidates differ from the inauguration speeches of presidents? How do preachers differ in their rhythmic styles? A web application for detecting and visualizing textual rhythm has straightforward teaching applications in linguistics, literature, music, political science, and religious studies, while also being useful for professional linguists interested in the analysis of rhythm in language.
We will develop three kinds of new technology based on earlier work with undergraduate and graduate student researchers:

For metrical parsing, we will build on the phonological software PROSODIC (Anttila, Heuser, and Falk 2010-2011, https://github.com/quadrismegistus/prosodic). PROSODIC has been used to detect verse scansion (Algee-Hewitt et al. 2015), but needs modification to handle sentence prosody. In particular, PROSODIC needs to be coupled with the Stanford Parser (Chen and Manning 2014, http://nlp.stanford.edu/software/lex-parser.shtml) that produces syntactic dependency trees essential for metrical parsing. PROSODIC allows the user to parse their own texts rhythmically.

For rhythm visualization, we will develop a web application that allows the user to drop text into PROSODIC, compute its rhythm, and visualize the result. In this way, the user can inspect the rhythmic profile of a text on the fly. A key person in the project will be the undergraduate web programmer whose job it will be to develop a user interface for PROSODIC.

For linguistic analysis, we will build on OTORDER (Djalali and Jeffers 2015, www.otorder.com), a web application for developing optimality-theoretic grammars (Prince and Smolensky 1993/2004). OTORDER allows the linguist to develop grammars capable of dealing with variation, but it is currently limited to fairly small inputs. Solving the computational problems involved with larger inputs is necessary for making the web application widely useful.

Through financial assistance from the School of Humanities and Sciences, we will have a dedicated server (otorder.stanford.edu) available at the Stanford Research Computing Facility (SRCF). The Denning grant is essential for hiring research personnel.